A STUDY OF THE NATURE OF DIFFICULTY

STANFORD UNIVERSITY LIBRARY
APR 1927

BY

JACOB S. ORLEANS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE FACULTY OF PHILOSOPHY, COLUMBIA UNIVERSITY.

STANFORD UNIVERSITY LIBRARIES
WITHDRAWN

BUREAU OF PUBLICATIONS
Teachers College, Columbia University
NEW YORK CITY
1926

81
20
CLASSED
Dissertation

A STUDY OF THE NATURE OF DIFFICULTY

BY

JACOB S. ORLEANS

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Faculty of Philosophy, Columbia University.

Bureau of Publications
Teachers College, Columbia University
New York City
1926

Copyright, 1926, by Teachers College, Columbia University

ACKNOWLEDGMENTS

This study could not have been made without the coöperation of several individuals, and it is a pleasure to the author to acknowledge his indebtedness to them. Drs. E. O. Bregman and E. Bailor, Miss Ella Woodyard, Miss L. Hunsicker, and Mr. John Tilton devoted much time and energy to the laborious task of making the judgments used in this study. Drs. R. Pintner, E. L. Thorndike, A. T. Wylie, and J. Slawson provided most of the original test records used. To the Commonwealth Fund the writer is indebted for certain of the test material provided by Dr. Thorndike.

To Dr. A. I. Gates and Dr. R. Pintner, the writer is indebted for guidance and advice in carrying on the investigation. To Dr. E. L. Thorndike is due not merely an appreciation of assistance in outlining and thinking through the problem, but all that background which is necessary to make such a study possible. Above all, the tireless assistance and encouragement of his wife has enabled the writer to carry the study through to completion.

J. S. O.

CONTENTS

CHAP.		PAGE
I.	Introduction	1
	The Problem	1
	Previous Studies of Difficulty	2
	Preliminary Statement	3
II.	Materials and Methods	5
	Tests Used and Subjects Tested	5
	Difficulty Values	6
	Characteristics and Judgments	9
III.	Reliability and Validity	15
	Difficulty Values	15
	Judgment Values	17
IV.	Statistical Treatment of Data	20
	Correlations	20
	Interpretation of Results	22
V.	Difficulty and Intelligence	29
VI.	Levels of Difficulty	33
VII.	Summary and Conclusions	36
	Bibliography	38

A STUDY OF THE NATURE OF DIFFICULTY

CHAPTER I

INTRODUCTION

THE PROBLEM

Both in analyzing learning processes and in mental and educational measurements, the difficulty of the subject matter is an important factor. The emphasis in learning is on that organization of bonds that can be formed with the least difficulty; and in the development of standardized testing, the attempt has been almost uniform to include elements in a test arranged in the order of difficulty.

It is somewhat surprising, in view of this, that an analysis of difficulty, leading to more efficient application of psychological and statistical methods to educational practice, has not been attempted. Such a study is obviously complex and laborious. Still, as will be shortly indicated, virtually no evidence has been found of an attempt to study the nature of difficulty. That the desirability of having more detailed knowledge of the degree of difficulty of tasks is appreciated, is evidenced by such statements as this: "The derivation of tests involves the evaluation of the elements of which the tests are composed. We need a far larger body of these evaluated elements than now exists. Indeed, there is ground for believing that if all the elements which are now used in the teaching process were evaluated, a much more intelligent use would be made of them. For example, if all the problems in an arithmetic had a difficulty rating, it is clear that fewer misunderstandings between teacher and pupils would arise, that more reasonable assignments would be made and that more satisfactory gradation and adaptation of material to the needs of the pupils would be possible." (Buckingham, 1921.)

The need here indicated for a more widespread evaluation of test elements is a real one and makes for greater efficiency in teaching and testing. But more fundamental in both these fields is the nature of difficulty—an understanding of what causes diffi-

culty. Such knowledge allows for an analysis of learning and how it functions. It is desirable to know what subject matter is suitable for different mental levels and different individuals. More desirable is knowledge of why such subject matter is suitable for such levels and such individuals.

This study is an indication of how an analysis of difficulty may be made, and is a first step in making it, to the end that learning and the measurement of learning may be more efficiently performed. Difficulty may be viewed from the standpoint of (1) the task responded to, the obviousness, complexity, and the like, of the task; (2) the neural medium allowing for the response to the situation, the complexity of the patterns involved, and the difficulty of their formation; (3) the response made. The present state of our neurological knowledge makes the second of these possibilities not feasible. The last gives merely a measure of the difficulties of the task. The first allows for an analysis of this difficulty. In other words, by studying the nature of the situation with relation to its difficulty, an understanding of the meaning of difficulty can be arrived at. This mode of procedure has been followed in making the present study.

In carrying out the analysis, an attempt has been made to answer these questions: What characteristics of subject matter make it difficult? To what extent are they factors in causing difficulty? Does the influence of these factors vary with the age and grade of the subjects tested? Does difficulty vary with intelligence? This study is limited to the answering of these questions.

PREVIOUS STUDIES OF DIFFICULTY

It is interesting to survey the available literature to discover what has been attempted in this field. A canvass was made of educational publications with this in view, covering the following journals:

British Journal of Psychology	Vol. 9–12	1917–1921
Educational Review	Vol. 23–30	1915–1922
Elementary School Journal	Vol. 16–22	1915–1921
Journal of Applied Psychology	Vol. 1– 5	1917–1921
Journal of Educational Psychology	Vol. 6–13	1915–1922
Journal of Educational Research	Vol. 1– 6	1920–1922
Journal of Experimental Pedagogy	Vol. 1– 8	1916–1922
Psychological Clinic	Vol. 11–13	1917–1919
Psychological Review	Vol. 28–29	1921–1922

School and Society	Vol. 1–15	1915–1922
School Review	Vol. 23–30	1915–1922
School Science and Mathematics	Vol. 18–22	1918–1922
Training School Bulletin	Vol. 14–19	1917–1922

Dr. Clara Chassell Cooper made a first survey, choosing for further study such articles as made any reference to difficulty. This list [1] on further examination showed a startling lack of objective data on the subject. The papers varied in this respect from mere mention, in some cases, that difficulty had been taken into account in constructing a test, to elaborate tables in a few articles, giving the difficulty of tasks for age or grade groups or both. Most of the papers were concerned with intelligence and educational tests. A few, however, were concerned with the learning processes involved in school subjects.[2] In only one article was an analysis of difficulty made. This was in a report of a pronunciation test made by Dr. A. I. Gates.[3]

PRELIMINARY STATEMENT

The difficulty of a situation to be responded to is defined objectively as the extent to which groups are able to make the correct response to the situation. Or, more specifically, the difficulty of a task for a given group is measured by the percent that the total correct response made by the group is of the total possible correct response. For example, if the score assigned to an element of a test is 3 and the total score achieved by a representative group of two hundred fifth grade pupils attempting this element is 450, then the difficulty of that element for fifth grade pupils is 75 percent. The higher the percent for a group, the easier the task for that group, and vice versa.

The first part of this study is devoted to a study of the difficulty of certain elements of standardized tests, an analysis of certain characteristics of these elements, a study of the relation of these characteristics to difficulty, and the differences in these relations for different age and grade groups. This is followed by a study of the relation between difficulty and intelligence, and a comparison of levels of difficulty.

[1] This list was obtained for the writer's use by the kindness of Dr. Cooper.

[2] Appendix I gives a list of the articles that offer objective data on the difficulty of test elements or other tasks.

[3] Gates, A. I., "A Test of Ability in Pronunciation of Words," *Teachers College Record*, Vol. 26, p. 205. (This article was discovered apart from the survey referred to above.)

To make such an investigation feasible it was thought desirable to have a variety of tasks for a large number of subjects in each of several groups of different status. In attempting to satisfy these conditions, test elements were utilized rather than situations more like classroom work, and only those tests were chosen that had already been given and scored for some other purpose, e.g., in the standardization of the tests. Treatment of the results and their interpretation lead to the opinion that a wiser procedure would have been to work with a limited number of test elements carefully selected, and a larger number of subjects, with a more specific rather than a general treatment of the data.

The test elements were all attempted under timed conditions. This factor must be kept in mind throughout, particularly in the interpretation of results. A time limit produces, of course, a different situation from one in which there is no time limit. The difference in situation may be due to (1) the form in which the task is put, certain types of tasks being more amenable to timed conditions than others; (2) the subject matter employed; (3) the attitude or set of the worker whose approach to the tasks must vary with the restraint under which he works. What might appear obvious to one working leisurely might, on the other hand, escape the attention of the same individual working hurriedly. An apparently elaborate situation may dissolve into a simple one when the subject is not hurried. If he is working under the stress of a time limit the apparent elaborateness may make the task confusing at the outset. The same may be true with other characteristics of the subject matter involved. It would be instructive to determine the influence of a time limit on difficulty of tasks. This could be done by determining the difficulty values under each of the two conditions and analyzing them.

CHAPTER II

MATERIALS AND METHODS

TESTS USED AND SUBJECTS TESTED

This section of the investigation makes use of the following standardized tests:

1. Pintner Non-Language Scale—Tests 4, 5, and 6.
2. Thorndike Non-Verbal Test—Tests 1 and 2, Forms 1A, 1F, 1G, 1K, 1L.
3. Thorndike Non-Verbal Test—Tests 1, 2, and 4, Forms IIA, IIB and IIL.
4. Wylie Opposites Test—Forms A, B, C, and D.
5. Trabue Completion-Test Language Scales—B, C, D, E, F, J, K, L, and M.

In these tests are included the following types of elements:

1. Picture completion.
2. Picture reorganization.
3. Geometrical completion.
4. Geometrical analogy.
5. Geometrical series.
6. Verbal opposites.
7. Sentence completion.
8. Arithmetic computation.

Table I shows the distribution of these elements among the above tests. There are in all 46 picture elements, 150 space elements, 220 verbal elements, and 60 number elements. This unequal distribution is due to the fact that only tests were used for this study that had already been given and scored. This was deemed necessary, as indicated above, to insure a sufficient variety of elements and a sufficient number of subjects to make the data reliable and the analysis valid and meaningful.

The Pintner Non-Language Scale [1] was administered in part in 1917 to 2,200 children in the public schools of Ohio in standardizing the test, and in part in 1922 to 2,450 children in the New York City public schools. The Thorndike Non-Verbal Test was given under the direction of Dr. John Slawson, of the New York State

[1] The original test blanks were available for the Pintner Non-Language Scale, the Thorndike Non-Verbal Test, and the Wylie Opposites Test. I am indebted to Drs. Thorndike, Pintner, Wylie, and Slawson for their kindness in making available to me the original material of their tests.

The data for the Trabue Completion Test were taken from Professor Trabue's thesis, *Completion-Test Language Scales* (Trabue, 1916), where they are given.

TABLE I

DISTRIBUTION OF THE SEVERAL TYPES OF TEST ELEMENTS IN THE TESTS FROM WHICH THEY WERE CHOSEN

Type of Test Element	Pintner Non-Verbal	Thorndike Non-Language *	Thorndike Non-Language †	Wylie Opposites	Trabue Completion	Total
Picture Completion.....	10		30			40
Picture Reorganization..	6					6
Geometrical Completion	10	50				60
Geometrical Analogy....			30			30
Geometrical Series.....		60				60
Verbal Opposites.....				140		140
Sentence Completion....					80	80
Arithmetic Computation			60			60

* Forms IA, IF, IG, IK, IL.
† Forms IIA, IIB, IIL.

Board of Charities, to about 1,500 boys, half of the male juvenile delinquent population in institutions in New York State (Slawson, 1924); and under the direction of Professor Thorndike to 3,300 elementary school, high school, and university students in New York City and Philadelphia in 1919, when standardizing the test. The Wylie Opposites Test was given to 2,500 elementary school, high school, and university students in New York City in 1921, when the test was standardized.

Table II shows the distribution of tests taken by different age and grade groups. In a number of instances, not included in the table, where the number of cases was small, they were not used for reasons of reliability to be discussed below.

DIFFICULTY VALUES

The scores obtained were tabulated for each age or grade group for each element attempted by that group. In some cases the

TABLE II

NUMBER OF SUBJECTS THAT TOOK EACH TEST, BY CHRONOLOGICAL AGES, MENTAL AGES, OR GRADES

Type of Test Element	Test	Chronological Ages							
		7	8	9	10	11	12	13	14
Picture Completion	Pintner 4	178	465	735	717	670	637	666	570
Picture Reorganization .	Pintner 6	178	465	735	717	670	637	666	570
Geometric Completion ..	Pintner 5	178	465	735	717	670	637	666	570
		Mental Ages							
		7	8	9	10	11	12	13	14
Picture Completion	Thorndike IIB 1	64	100	134	238	252	250	120	101
Geometric Completion ..	Thorndike								
	IA 2	64	100	134	238	234	250	120	101
	IL 2	65	100	134	231	252	249	120	101
Geometric Analogy	Thorndike IIB 2	64	100	134	238	252	250	120	101
Geometric Series	Thorndike								
	IA 1	64	100	134	238	234	250	120	101
	IL 1	65	100	134	231	252	249	120	101
Arithmetic Computation	Thorndike IIB 4	64	100	134	238	252	250	120	101

		Grades									Adult *
		4	5	6	7	8	9	10	11	12	
Picture Analogy	Thorndike										
	IIA 1			102						238	
	IIB 1			158		100					
	IIL 1			177						126	
Geometric Completion ..	Thorndike										
	IA 2			161		151				179	131
	IF 2			246	116	143					
	IG 2			209	76	129					
	IK 2			254							
	IL 2			303		93				181	144

* The adult group is composed of graduate students of education.

TABLE II (*Continued*)

Type of Test Element	Test	Grades									Adult *
		4	5	6	7	8	9	10	11	12	
Geometric Analogy	Thorndike										
	IIA 2			102						238	
	IIB 2			158		100					
	IIL 2			177						126	
Geometric Series	Thorndike										
	IA 1			161		151				179	131
	IF 1			246	116	143					
	IG 1			209	76	129					
	IK 1			254							
	IL 1			303	93					181	144
Verbal Opposites	Wylie A	218	419	447	455	408	70	172			407
	B	220	421	387	456	417	73	160			427
	C	216	417	418	454	416	70	166			363
	D	219	421	414	455	413	70	173			430
Arithmetic Computation	Thorndike										
	IIA 4			102							238
	IIB 4			158	100						
	IIL 4			177							126
Verbal (Sentence) Completions	Trabue Forms B, C, D, E, F, J, K, L, M.	Populations varied from 114 to 2646 for the different elements.†									

† The data for the Trabue Completion Sentences were obtained from the giving of the scale before the sentences were grouped into the several forms. Therefore the populations within each grade group vary for the several sentences in any form.

ages of the subjects were given on the test blanks. In other cases the grade status alone was given. From Table II may be gathered the basis for the grouping in the case of each test.

The total score obtained by each age or grade group on any element, divided by the total obtainable score for that group, gives the difficulty value for that element for that group. A difficulty value means, then, the percent of correct response. Where one of several values may be assigned for the responses to a task, no one individual may make the correct response, and still the difficulty value may be fairly high if most of the subjects make fairly good responses.

To obtain these difficulty values, the score on each element of each of the tests was tabulated by chronological age, mental age, or grade as the case may have been. These scores were totalled for each group for each element and this total divided by the total obtainable score. Table III [1] gives these values for each of the 476 elements for each group of sufficient size that attempted the various tasks.

CHARACTERISTICS AND JUDGMENTS

The next step is to determine what characteristics or traits of these elements need be studied with reference to difficulty, and then to evaluate these traits. There is no method for determining *a priori* just which traits to study. One can only choose such as would seem to have some relation to difficulty and then determine the extent to which this relation exists. The characteristics chosen are:

Complexity	Elaborateness
Abstractness	Symbolicness
Obviousness	Familiarity

Intrinsic Interest

Complexity refers to how involved the task is in its performance; elaborateness to how involved the situation appears. The former refers to the content or process of working out the solution; the latter refers to the form in which the task is set. Abstractness refers to the lack of concreteness in the performance or process of working out the task, while symbolicness refers to the lack of concreteness of representation of the situation. Obviousness refers to acquaintance with the performance of the task; familiarity refers to acquaintance with the form in which it is represented. Interest refers to the influence of the task in commanding the attention and further activity of the worker.

It may appear that complexity and elaborateness are synonymous, as may be abstractness and symbolicness, familiarity and obviousness. However, complexity, abstractness, and obviousness are defined as referring to the process of responding to the situation or performing the task; while elaborateness, symbolic-

[1] Only part of this table is shown here. The complete table, together with the complete forms of Tables IV, XIV, XV, and XVI, is deposited in the Library at Teachers College, Columbia University.

TABLE III

(Only the first part of this table is shown here)

DIFFICULTY VALUES OF THE 474 TEST ELEMENTS BY GRADES AND AGE GROUPS

Test Elements		Mental Ages									Grades							
		7	8	9	10	11	12	13	14	Ad	4	5	6	7	8	9	10	12
Thorndike Non-Verbal																		
Form IA	1	28	37	42	59	50	77	82	86	87			86		83			95
Test 1	2	11	20	27	32	34	54	63	64	80			71		66			91
	3	9	19	21	33	41	58	61	68	80			71		65			92
	4	23	35	46	58	59	79	78	85	90			81		78			96
	5	13	24	31	39	44	64	68	70	84			72		63			93
	6	25	31	31	54	56	76	80	83	91			71		65			96
	7	9	16	15	29	33	63	59	65	77			55		54			90
	8	6	17	23	35	39	60	66	69	79			53		46			91
	9	2	7	18	23	43	46	51	57	68			34		30			66
	10	3	9	16	26	33	52	57	62	78			35		28			67
	11	0	5	4	8	14	16	26	35	48			21		17			48
	12	3	7	6	19	23	35	41	47	70			28		17			50
Form IF	1												66	60	66			
Test 1	2												27	37	44			
	3												65	45	66			
	4												40	49	57			
	5												61	62	69			
	6												44	52	60			
	7												54	58	62			
	8												59	56	65			
	9												49	59	64			
	10												32	47	50			
	11												30	51	53			
	12												13	22	27			
Form IG	1												70	83	74			
Test 1	2												68	76	76			
	3												53	63	64			
	4												61	72	70			
	5												53	64	67			
	6												47	50	53			
	7												28	29	40			
	8												31	33	49			
	9												33	29	50			
	10												30	35	42			
	11												24	23	33			
	12												16	16	26			
Form IK	1												86					
Test 1	2												58					
	3												62					
	4												80					
	5												66					
	6												69					
	7												60					
	8												46					
	9												38					
	10												52					
	11												54					
	12												53					

ness, and familiarity are defined as referring to the form in which the situation is presented. Hence, while a situation may be elaborate in structure, it may be simple in the process of working it out. Or, a task may be elaborate and at the same time either concrete or symbolic in structure. Or, it may be abstract in the process of working it, but either concrete or symbolic in structure; for an expert stenographer, reading shorthand is a more abstract process than typing shorthand. Reading shorthand may be, to an individual, no more abstract than solving a rebus. But these two situations differ greatly in the symbolicness of their presentation.

It is difficult to obtain objective measures of these traits for most tasks, particularly for the types of situations employed in this study. The elaborateness of words may be measured in terms of their length and number of syllables. The elaborateness of sentence completions by the length of the sentences, the number of completions, and the like. But it is, perhaps, impossible to obtain an accurate measure of the complexity of the process involved in finding the opposite of a word, or in completing a series of dots and circles. These traits are mental attributes that are difficult to express accurately. To evaluate them it was decided to obtain the judgments of individuals competent to pass on them.

The seven characteristics chosen—complexity, elaborateness, abstractness, symbolicness, familiarity, obviousness, and intrinsic interest—may be expressed quantitatively, as will be shortly indicated. Tasks vary from zero complexity to maximum complexity. A number of situations may be arranged in a series from "having least" to "having most" of each of these traits.

The judgments of the traits mentioned above totalled 3,332 in all, seven for each of the 476 elements involved. In view of the extent of this task, the question of the most desirable number of judges could hardly be considered. For each trait of each element, five judgments were obtained. The judges included four research assistants in the Institute of Educational Research of Teachers College, and an assistant in the department of educational psychology of the same institution. All five judges had had experience in making such judgments and were well fitted for the work. Since the writer was acquainted with the difficulty values of the elements, it was thought unwise for him to participate

in the work, inasmuch as his judgments would have been biased thereby. Each judge was supplied with a set of the test elements, each element on a separate paper; a sheet of quadri-ruled paper 20 inches long, divided into one-inch squares, each of these squares subdivided into squares one-tenth of an inch to the side; and the instructions on the following page.

It might have been wiser to have chosen tasks representative of zero and of 100 for each of the traits. It was felt, however, that this was part of the judgment, and that it might confuse those judges who did not agree with the choice made. It was left to each judge to evaluate or represent these reference points as he or she saw fit.

INSTRUCTIONS TO JUDGES

The accompanying envelope contains

		No. of Elements
1. Thorndike Non-Verbal Test		
Forms	Test 1—geometrical series	60
1A, 1F, 1G, 1K, 1L	Test 2—geometrical completion	50
2A, 2B, 2L	Test 1—picture completion	30
	Test 2—geometrical analogy	30
	Test 4—arithmetic computation	60
2. Pintner Non-Language Test		
	Test 4—picture completion	10
	Test 5—geometrical completion	10
	Test 6—picture rearrangement	6
3. Wylie Opposites	Forms A, B, C, D	140
4. Trabue Completions	Forms B, C, D, E, F, J, K, L, M	80
	Total	476

The back of each element is labeled with its form, test and element number. Thus 2L²4 means Form 2L, Test 2, Element 4. TB6 means Trabue, Form B, Element 6.

These 476 elements are to be rated on each of the following characteristics:

1. complexity
2. elaborateness
3. abstractness
4. symbolicness
5. familiarity
6. obviousness
7. intrinsic interest

1, 3, and 6 refer to the process of working each element; 2, 4, and 5 refer to the structure and subject matter.

The following procedure is suggested in making the judgments:

1. On the accompanying graph paper, choose a 20-inch line and mark the left end 0 and the right end 100.
2. Choose one set of test elements (e.g., geometrical series) and arrange the elements in the order of their complexity.
3. Look at the least complex of these elements. Indicate on the 20-inch line how far above zero you estimate it to be.
4. Look at the most complex of these elements. Indicate on the line how far below 100 you estimate it to be.
5. Locate the position of the other elements between these limits.
6. Record on the record sheets provided the values for these elements in the column headed complexity. Each large square on the graph paper counts 5; each small square counts .5.
7. Choose another set of elements and follow the same procedure. Continue with the rest of the types of elements.
8. Repeat the same procedure, choosing another line on the graph paper for elaborateness, marking its ends 0 and 100.
9. Repeat the same procedure for the other characteristics.

You may, if you wish, place the elements of one set in several groups instead of in a series from lowest to highest on any one characteristic. If you judge all the elements in each of these groups to be of equal value in the characteristic, assign them that value. But if you judge them to differ within the group, then give separate values to the elements of the group.

If you prefer, in the case of any characteristic, to rank the 476 elements as one group instead of as several different sets, you may do so.

You may, if you wish, rank two or more sets at once, or use any other method.

This method of making the evaluation does more than merely arrange the items in rank order. It attempts to give them relative positions in terms of quantitative measures expressed in terms of the percentage of the distance between the least and most in each trait. It does away, moreover, with the difficulty of having to rate on the same basis such widely different tasks as picture rearrangement and arithmetic computation.

The judgments given were recorded after each element and the median computed for each trait for each element. These medians are given in Table IV.[1]

[1] See footnote, page 9.

TABLE IV

(Only the first page of this table is shown here)

MEDIAN OF FIVE JUDGMENTS FOR EACH OF THE CHARACTERISTICS—COMPLEXITY, ELABORATENESS, ABSTRACTNESS, SYMBOLICNESS, FAMILIARITY, OBVIOUSNESS, AND INTEREST, FOR EACH TEST ELEMENT

Test Elements		Characteristics							Test Elements		Characteristics						
		C	E	A	S	F	O	I			C	E	A	S	F	O	I
Thorndike Form IA Test 1	1	10	15	10	15	75	93	15	Thorndike Form IL Test 1	1	5	5	10	10	81	95	10
	2	20	25	10	18	76	85	14		2	7	7	10	10	74	99	10
	3	15	20	10	15	76	88	18		3	12	12	10	10	74	88	10
	4	35	35	10	18	76	65	18		4	14	14	10	10	74	86	10
	5	25	30	10	18	76	75	18		5	22	30	10	10	74	75	10
	6	29	40	10	18	76	50	18		6	22	35	10	10	74	75	10
	7	40	55	10	18	75	38	20		7	20	12	10	18	71	83	19
	8	39	40	10	15	70	60	13		8	25	33	10	10	74	45	10
	9	30	8	10	10	79	73	12		9	29	20	10	15	76	70	18
	10	25	39	10	10	74	60	10		10	20	18	10	15	71	80	19
	11	25	39	10	10	74	58	10		11	29	40	10	15	76	60	18
	12	37	50	10	18	74	45	12		12	29	35	10	18	76	65	18
Form IF Test 1	1	10	15	10	10	74	90	10	Form IA Test 2	1	12	10	25	10	74	90	30
	2	15	15	10	15	75	85	15		2	12	10	25	10	74	90	30
	3	20	20	10	10	74	80	10		3	15	15	25	10	74	85	30
	4	25	25	10	18	76	75	18		4	15	15	25	10	74	70	30
	5	25	34	10	10	74	50	10		5	20	20	25	10	74	80	30
	6	22	35	10	10	74	75	10		6	25	20	25	10	74	70	30
	7	30	40	10	18	71	70	20		7	34	25	25	10	74	60	30
	8	20	28	10	10	71	65	10		8	50	24	25	10	74	50	30
	9	25	35	10	10	74	45	10		9	21	20	25	10	74	65	30
	10	20	30	10	15	71	70	10		10	50	24	25	10	74	50	30
	11	30	35	10	18	76	70	20									
	12	25	40	10	10	74	48	10	Form IF Test 2	1	10	15	25	10	72	90	30
										2	10	5	25	10	72	75	30
Form IG Test 1	1	10	15	10	10	76	90	14		3	13	15	25	10	72	85	30
	2	15	20	10	18	71	93	16		4	24	15	25	10	72	55	30
	3	30	35	10	18	71	78	16		5	16	20	25	10	72	70	30
	4	20	30	10	18	71	80	16									
	5	37	43	10	18	68	75	13		6	20	21	25	10	53	80	30
	6	25	20	10	15	65	75	17		7	45	25	25	10	53	50	30
										8	23	30	25	10	53	55	30
	7	35	35	10	15	65	65	17		9	35	24	25	10	53	25	30
	8	35	40	10	15	65	70	17		10	47	27	25	10	53	40	30
	9	25	36	10	10	74	55	10									
	10	20	25	10	10	71	88	10	Form IG Test 2	1	13	15	25	10	75	85	30
	11	20	36	10	10	71	65	10		2	12	15	25	10	75	80	30
	12	37	55	10	18	70	55	20		3	30	25	25	10	75	70	30
										4	30	21	25	10	75	75	30
Form IK Test 1	1	10	15	10	10	76	90	14		5	28	20	25	10	75	55	30
	2	20	20	10	15	76	88	18									
	3	20	20	10	10	70	18	15		6	17	20	25	10	75	85	30
	4	15	15	10	10	70	75	15		7	18	24	25	10	75	55	30
	5	35	12	10	10	74	73	12		8	48	27	25	10	75	35	30
	6	40	40	10	15	71	60	20		9	48	25	25	10	75	45	30
										10	49	27	25	10	75	40	30
	7	20	25	10	10	76	85	14									
	8	30	12	10	10	72	73	12	Form IK Test 2	1	20	10	25	10	60	80	30
	9	22	38	10	10	74	55	10		2	12	5	25	10	60	87	30
	10	25	39	10	10	74	50	10		3	12	5	25	10	65	90	30
	11	40	45	10	18	75	45	20		4	25	25	25	10	65	50	30
	12	37	50	10	18	71	35	20		5	14	5	25	10	60	75	30

CHAPTER III

RELIABILITY AND VALIDITY OF MEASURES

DIFFICULTY VALUES

Before proceeding with the analysis of difficulty, it is necessary to determine that the data are sufficiently accurate and valid for the purpose, or to determine what inaccuracies exist, so that they may be allowed for in the treatment and interpretation of the data. The accuracy or reliability of the difficulty measures is given by the probable errors of the percents in Table IV. The probable error of a percent is given by the formula (Kelley, 1923):

$$\text{P.E.}_{\text{percent}} = .6745\sqrt{\frac{pq}{n}}$$

where p = the percent
$q = 100 - p$
n = the number of cases from which p is derived.

Table VI gives the probable errors of percents from 1 to 50 for different values of n. Beyond 50% the table is repeated. $\text{P.E.}_{51\%} = \text{P.E.}_{49\%}$; $\text{P.E.}_{52\%} = \text{P.E.}_{48\%}$; $\text{P.E.}_{72\%} = \text{P.E.}_{28\%}$, etc. This is obvious from the formula, since $p + q = 100\%$. To determine the probable errors of the percents in Table III, the populations on the basis of which they were computed must be known. These are given in Table II. It will be seen from this table that in all but six cases the population is at least 100. In those six instances the population is 65 or over. In the statistical treatment of the data, the measures were always grouped within a range of at least 5% and at times within a range of 10%. This is done not simply as a means to lessen the statistical labor involved, but rather because, by the very definition of difficulty, it is needless to draw sharp distinctions between tasks differing but slightly in difficulty. With this in mind, a population of 100 is considered large enough to give a sufficiently reliable measure.

The validity of the percents which are used as difficulty measures depends on the selection of the subjects. If the subjects are unselected groups, then the difficulty values measure what they

TABLE V

PROBABLE ERRORS FOR PERCENTS OF CORRECT RESPONSE ON ANY TEST ELEMENT FOR POPULATIONS OF DIFFERENT SIZES

FROM THE FORMULA: $P.E. = .6745\sqrt{\frac{pq}{n}}$

Percent	Populations														
	50	75	100	125	150	175	200	225	250	275	300	350	400	450	500
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.0	0.8	0.7	0.6	0.6	0.5	0.5	0.5	0.4	0.4	0.4	0.4	0.3	0.3	0.3
2	1.3	1.1	0.9	0.8	0.8	0.8	0.7	0.7	0.6	0.6	0.6	0.6	0.5	0.5	0.4
3	1.6	1.3	1.2	1.0	0.9	0.9	0.8	0.8	0.7	0.7	0.7	0.6	0.6	0.5	0.5
4	1.9	1.5	1.3	1.2	1.1	1.0	0.9	0.9	0.8	0.8	0.8	0.7	0.7	0.6	0.6
5	2.1	1.7	1.5	1.3	1.2	1.1	1.0	1.0	0.9	0.9	0.9	0.8	0.7	0.7	0.7
6	2.3	1.9	1.6	1.4	1.3	1.2	1.1	1.1	1.0	1.0	0.9	0.9	0.8	0.8	0.7
7	2.4	2.0	1.7	1.5	1.4	1.3	1.2	1.2	1.1	1.0	1.0	0.9	0.9	0.8	0.8
8	2.6	2.1	1.8	1.6	1.5	1.4	1.3	1.2	1.2	1.1	1.1	1.0	0.9	0.9	0.8
9	2.7	2.2	1.9	1.7	1.6	1.5	1.4	1.3	1.2	1.2	1.1	1.0	1.0	0.9	0.9
10	2.9	2.3	2.0	1.8	1.7	1.5	1.4	1.4	1.3	1.2	1.2	1.1	1.0	1.0	0.9
11	3.0	2.4	2.1	1.9	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.1	1.1	1.0	1.0
12	3.1	2.5	2.2	2.0	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.2	1.1	1.0	1.0
13	3.2	2.6	2.3	2.0	1.9	1.7	1.6	1.5	1.4	1.4	1.3	1.2	1.1	1.1	1.0
14	3.3	2.7	2.3	2.1	1.9	1.8	1.7	1.5	1.5	1.4	1.4	1.2	1.2	1.1	1.1
15	3.4	2.8	2.4	2.2	2.0	1.8	1.7	1.6	1.5	1.5	1.4	1.3	1.2	1.1	1.1
16	3.5	2.9	2.5	2.2	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.2	1.1
17	3.6	2.9	2.5	2.3	2.1	1.9	1.8	1.7	1.6	1.5	1.5	1.4	1.3	1.2	1.1
18	3.7	3.0	2.6	2.3	2.1	2.0	1.8	1.7	1.6	1.6	1.5	1.4	1.3	1.2	1.2
19	3.7	3.1	2.7	2.4	2.2	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.3	1.2
20	3.8	3.1	2.7	2.4	2.2	2.0	1.9	1.8	1.7	1.6	1.6	1.4	1.4	1.3	1.2
21	3.9	3.2	2.8	2.5	2.3	2.1	2.0	1.8	1.7	1.7	1.6	1.5	1.4	1.3	1.2
22	4.0	3.2	2.8	2.5	2.3	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.3
23	4.0	3.3	2.8	2.5	2.3	2.2	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.3
24	4.1	3.3	2.9	2.6	2.4	2.2	2.0	1.9	1.8	1.7	1.7	1.5	1.4	1.4	1.3
25	4.1	3.4	2.9	2.6	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3
26	4.2	3.4	3.0	2.6	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3
27	4.2	3.5	3.0	2.7	2.5	2.3	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3
28	4.3	3.5	3.0	2.7	2.5	2.3	2.1	2.0	1.9	1.8	1.8	1.6	1.5	1.4	1.4
29	4.3	3.5	3.1	2.7	2.5	2.3	2.2	2.0	1.9	1.9	1.8	1.6	1.5	1.4	1.4
30	4.4	3.6	3.1	2.8	2.5	2.3	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4
31	4.4	3.6	3.1	2.8	2.6	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4
32	4.5	3.6	3.2	2.8	2.6	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4
33	4.5	3.7	3.2	2.8	2.6	2.4	2.3	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4
34	4.5	3.7	3.2	2.9	2.6	2.4	2.3	2.1	2.0	1.9	1.9	1.7	1.6	1.5	1.4
35	4.6	3.7	3.2	2.9	2.6	2.4	2.3	2.1	2.0	1.9	1.9	1.7	1.6	1.5	1.4
36	4.6	3.7	3.2	2.9	2.6	2.5	2.3	2.2	2.1	2.0	1.9	1.7	1.6	1.5	1.5
37	4.6	3.8	3.3	2.9	2.7	2.5	2.3	2.2	2.1	2.0	1.9	1.7	1.6	1.5	1.5
38	4.6	3.8	3.3	2.9	2.7	2.5	2.3	2.2	2.1	2.0	1.9	1.7	1.6	1.5	1.5
39	4.7	3.8	3.3	2.9	2.7	2.5	2.3	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5
40	4.7	3.8	3.3	3.0	2.7	2.5	2.3	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5
41	4.7	3.8	3.3	3.0	2.7	2.5	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5
42	4.7	3.8	3.3	3.0	2.7	2.5	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5
43	4.7	3.9	3.3	3.0	2.7	2.5	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5
44	4.7	3.9	3.4	3.0	2.7	2.5	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5
45	4.8	3.9	3.4	3.0	2.7	2.5	2.4	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5
46	4.8	3.9	3.4	3.0	2.8	2.5	2.4	2.2	2.1	2.0	2.0	1.8	1.7	1.6	1.5
47	4.8	3.9	3.4	3.0	2.8	2.5	2.4	2.2	2.1	2.0	2.0	1.8	1.7	1.6	1.5
48	4.8	3.9	3.4	3.0	2.8	2.6	2.4	2.3	2.1	2.0	2.0	1.8	1.7	1.6	1.5
49	4.8	3.9	3.4	3.0	2.8	2.6	2.4	2.3	2.1	2.0	2.0	1.8	1.7	1.6	1.5
50	4.8	3.9	3.4	3.0	2.8	2.6	2.4	2.3	2.1	2.0	2.0	1.8	1.7	1.6	1.5

are supposed to measure—the extent to which a given group can respond correctly to given situations. It need only be pointed out that the subjects included in the chronological age groups and grade groups were chosen in the standardization of the tests involved, by experts in the field of educational and mental tests. Unselected groups are necessary for this purpose. The mental age groups are highly selected, and are meant to be, the average IQ of these groups being 80. They allow for a comparison between normal and subnormal groups. As a subnormal group the mental age groups can surely be considered as unselected!

JUDGMENT VALUES

The reliability of the judgments is determinable by having the same judges do at least part of their work over again, and by correlating the median of these judgments with the medians of those first made. It was not possible to do this in this study. No further statement can be made on this point beyond repeating that the judges were experts.

The validity of the judgments can be measured by the correlation between the judgments obtained and those of another group of judges. In lieu of this may be substituted the correlation between the judgments of two of the five judges and those of the other three. These correlations for each characteristic are given in Table VI. By means of the Spearman-Brown formula:

$$r_{aa} = \frac{ar}{1 - (a - 1)r}$$

r_{aa} = correlations between a number of judgments and the same number of other judgments,

r = correlation between one judgment and one other judgment,

a = the number of judgments,

the coefficients in Table VI are "corrected" to those in Table VIA, which accord more nearly with the present case.[1]

[1] Since an uneven number of judgments is used in this instance the correction was made from 2½ judgments to 5 judgments, the 2½ being taken as roughly representing a compromise between 2 and 3. Although this is not accurate, it is sufficient for the present purpose in indicating the validity of the judgments.

TABLE VI

CORRELATIONS OF TWO JUDGMENTS WITH THREE JUDGMENTS FOR EACH CHARACTERISTIC

Complexity	.494
Elaborateness	.533
Abstractness	.603
Symbolicness	.423
Familiarity	.611
Obviousness	.606
Interest	.077

TABLE VIA

CORRELATIONS OF TABLE VII MODIFIED TO REPRESENT THEORETICAL CORRELATIONS OF FIVE JUDGMENTS WITH FIVE JUDGMENTS

Complexity	.661
Elaborateness	.695
Abstractness	.752
Symbolicness	.595
Familiarity	.759
Obviousness	.755
Interest	.142

How high a coefficient is necessary or advisable for the data to be considered valid? The coefficient of alienation (Kelley, 1923)

$$k = \sqrt{1 - r^2}$$

aids us in determining this. If a correlation coefficient is computed between two traits, and a measure in one trait is known, the corresponding measure in the other trait can be estimated. The extent to which the error of this estimate or prediction differs from the error of a random estimate is the coefficient of alienation of the correlation between the traits in question. The formula gives the following table:

when	$r = .00$	$k = 1.00$
	$r = .10$	$k = .995$
	$r = .50$	$k = .87$
	$r = .80$	$k = .60$
	$r = .90$	$k = .44$
	$r = .95$	$k = .31$
	$r = .99$	$k = .14$
	$r = 1.00$	$k = .00$

In other words, when the correlation is zero, a prediction based on

the correlation is no better than a random guess. When the correlation is .50, the prediction based on it is still probably 7/8 as erroneous as a random guess. When the correlation is .99, the prediction based on it still has a probable error 1/7 as large as that of a random estimate.

This is true if the prediction made is required to be exact. But if it is to be made within a given range the probable error is lessened. As already stated, it is useless to draw fine distinctions between the difficulty of tasks. Hence, in the present instance the formula $k = \sqrt{1 - r^2}$ needs to be modified to accord with predictions within a range of, say, 5 or 10 percent, that is, from $.5\sigma$ to 1σ of the entire range.

It is possible that the difference in reliability of the estimates of the several traits may influence the magnitudes of the correlations based thereon. If such errors are random, correction for attenuation would eliminate their influence. That such errors, if they exist, are random is doubtful. A judge is rather inclined to be too severe or too lenient with all the judgments of one trait rather than to tend to compensate severity with leniency in judging. Such errors in estimates would not affect the correlations. The correction for attenuation should be made only when it is certain that the errors are random.

CHAPTER IV

STATISTICAL TREATMENT OF DATA

CORRELATIONS

The relation between the seven traits listed in Chapter II on the one hand, and difficulty on the other hand, can be determined from the correlations between the measures of these traits and the difficulty values. The correlations were computed from the formula:[1]

$$r = \frac{\sigma_x}{2\sigma_y} + \frac{\sigma_y}{2\sigma_x} - \frac{\frac{\Sigma D^2}{n} - (M_x - M_y)^2}{2\sigma_x\sigma_y}$$

in which σ_x = the standard deviation of the X distribution
σ_y = the standard deviation of the Y distribution
D = the difference between the X and Y scores for any individual
M_x = the mean of the X distribution
M_y = the mean of the Y distribution
n = the number of cases.

The correlations are given in Table VII. In interpreting this table it will be well to keep in mind (1) the discussion of the validity of the judgments given in the preceding chapter; (2) the reliability of the difficulty values; (3) the number of items in each correlation. The coefficients are not all strictly comparable, particularly for the last reason, viz: those for the chronological age groups are based on 26 test elements of three types, all non-verbal material; those for the grade groups and for the adults on over 150 elements, both verbal and non-verbal; those for the mental age groups on 84 test elements, all non-verbal, of 5 different types.

The difficulty values of Table III are given in terms of percent of correct response to accord with uniform practice. The correlations of the traits with these values should, then, be the opposites of the correlations with difficulty. The signs of the correlations have therefore been reversed. To avoid this modification,

[1] This is a variation of the usual Pearson formula. Its use does not involve plotting. In working several correlations involving the same variable much of the work is avoided by its use.

it would have been possible to use the opposites of the traits employed—simplicity for complexity, etc. This would have led to confusion in terminology, inasmuch as the same term is used for the converse of both complexity and elaborateness, and

TABLE VII

CORRELATIONS BETWEEN DIFFICULTY AND EACH OF SEVEN CHARACTERISTICS OF TEST ELEMENTS, BY AGE AND GRADE GROUPS

Groups	Characteristics						
	Comp.	Elab.	Abstr.	Symb.	Fam.	Obv.	Interest
Grade							
4	+.678	+.320	+.549	−.349	−.782	−.804	+.561
5	+.648	+.246	+.587	−.384	−.696	−.721	+.659
6	+.791	+.225	+.307	−.167	−.572	−.565	+.365
7	+.565	+.143	+.498	−.296	−.716	−.675	+.597
8	+.611	+.175	+.426	−.174	−.645	−.603	+.445
9	+.584	+.200	+.650	−.321	−.800	−.823	+.636
10	+.579	+.133	+.574	−.389	−.811	−.806	+.635
12	+.638	+.597	+.023	−.038	−.485	−.476	+.165
C. A.							
7	+.506	+.423	+.080	−.400	−.532	−.430	−.349
8	+.538	+.448	+.018	−.391	−.581	−.458	−.387
9	+.541	+.437	−.098	−.392	−.583	−.512	−.396
10	+.546	+.451	−.074	−.385	−.632	−.497	−.413
11	+.570	+.508	−.083	−.338	−.658	−.609	−.384
12	+.570	+.481	−.099	−.352	−.657	−.580	−.380
13	+.593	+.550	−.070	−.335	−.660	−.593	−.392
14	+.582	+.539	−.107	−.366	−.658	−.591	−.397
Adult	+.487	+.042	+.529	−.057	−.626	−.800	+.636
M. A.							
7	+.666	+.373	−.093	−.172	−.305	−.329	−.155
8	+.609	+.364	−.234	−.380	−.399	−.501	−.027
9	+.646	+.381	−.145	−.266	−.451	−.480	−.015
10	+.694	+.381	−.227	−.285	−.469	−.460	+.093
11	+.729	+.386	−.064	−.210	−.414	−.470	+.094
12	+.726	+.563	−.024	−.233	−.468	−.382	+.165
13	+.779	+.358	−.177	−.244	−.549	−.458	+.139
14	+.685	+.302	−.154	−.214	−.441	−.444	+.179

similarly for the converse of abstractness and symbolicness. The less confusing form is therefore chosen in reversing the signs of the correlations in Table VII, so that this table gives the correlations of the several traits with difficulty of response instead of with ease of response.

INTERPRETATION OF RESULTS

To clarify the table of coefficients, Table VIII is added, showing the distribution of correlations for each characteristic with difficulty. These two tables may be summarized as follows: (1) The different groups vary appreciably in the correlations of any trait with difficulty. (2) On the whole, complexity is most highly correlated with difficulty of such tasks as this study includes (median $r_{dc} = .609$); familiarity and obviousness are next in importance in determining difficulty (median $r_{df} = -.583$; $r_{do} = -.512$); elaborateness and symbolicness are correlated to only a slight degree with difficulty (median $r_{de} = .381$; $r_{ds} = -.321$); while abstractness and intrinsic interest seem to play no part in producing difficulty (median $r_{da} = -.064$; $r_{di} = .165$). (3) The

TABLE VIII

DISTRIBUTIONS OF THE CORRELATIONS IN TABLE VII

r	Characteristics						
	C	E	A	S	F	O	I
−.800					2	4	
−.700					2	1	
−.600					8	3	
−.500					5	6	
−.400				1	6	9	1
−.300				12	2	2	7
−.200			2	7			
−.100			4	3			1
−.000			8	2			2
.000		1	3				2
.100		3					4
.200		3					
.300		8	1				1
.400	1	5	2				1
.500	11	5	4				2
.600	9		1				4
.700	4						
Median	.609	.381	−.064	−.321	−.583	−.512	.165
Highest	.779	.597	.587	−.038	−.305	−.329	.659
Lowest	.487	.042	−.234	−.400	−.811	−.823	−.413

relations of the several traits to difficulty are different for the age groups, the grade groups, the mental age groups, and the adult group.

Such a general statement is hardly useful, unless it is also true in specific applications. That it is only a general statement can be inferred by an examination, for example, of the correlations for the adult group (Table VII). This group was composed of graduate students of education. For this group, abstractness and intrinsic interest are important factors in causing difficulty. Though superior adults are accustomed to abstract thinking, it appears from the correlation ($r_{da} = .529$) that the influence of abstractness on difficulty is not thereby entirely neutralized. From the above statement, it may be implied that, in general, the interesting nature of a task has no bearing on making it easy, or vice versa. For the adult group the more interesting tasks are, on the whole, the more difficult ones. However, this may be an accompaniment and not a cause. It may be nearer the truth to say that for superior adults the more difficult the task, the more interesting it is, for the subject matter used here. With more difficult tasks included, this might not be the case.

It would appear, a priori, that, since superior adults are accustomed to working with symbols to a marked degree, the symbolicness of such tasks as are here dealt with would hardly affect their difficulty. Such is the case ($r_{ds} = .057$ for adults). Similarly, the elaborateness of the situation is a matter of indifference to the adults in this group. Apparently complexity of the situation is a factor in making for difficulty with adults ($r_{dc} = .487$ for adults). But the largest correlations for this group are for familiarity and obviousness ($r_{df} = .626$; $r_{do} = .800$). This must mean that strangeness of situations and the inability to get the solution quickly are the main causes or indications of difficulty with this group.

It would be worthwhile to analyze further the relation of difficulty to the several characteristics for the adult group, by means of partial correlations. The data presented, however, are not sufficiently reliable to give meaningful partial correlations.

The outstanding discrepancies between the correlations given in Table VIII are:

1. Correlations between interest and difficulty are negative for the grade groups; positive for the C.A.* groups; zero for the M.A.* groups.

* C.A. groups means those grouped according to chronological age.
M.A. groups means those grouped according to mental age.

2. Correlations between difficulty and abstractness are negative for the grade groups; zero for the C.A. groups; slightly positive for the M.A. groups.
3. Correlations between difficulty and abstractness are much lower for the grade groups than for either the C.A. or the M.A. groups.

These discrepancies may be due in part to the extent to which the data lack reliability; in part to difference in number of cases (26 elements for the C.A., 84 for the M.A., and 150 or more for the grade groups); in part to the differences between the test elements for each group on which the correlations are based; and in part to the validity of the difficulty and judgment values for the groups. Table IX shows this difference between the test elements for each group. These discrepancies are inherent in the method of obtaining the data and are unavoidable, if enough cases and a diversity of material are to be studied under limited conditions. They may also be inherent in the nature of the judgments; that is, since the judgments were made by adults, certain factors might have been stressed and others disregarded. A larger number of cases, more judges, greater uniformity in the types of tasks to which the several groups are subjected would perhaps overcome in large part the discrepancies listed.

The differences in constitution between the grade and C.A. groups would hardly produce such discrepancies. They would probably tend only to restrict the correlations more in one case than in the other. The highly selected nature of the M.A. groups may in large part be the cause of such differences as appear between the correlations for this group and for the other two.

No attempt was made to limit the viewpoint of the judges. The judgments made are therefore from an adult point of view. This would probably affect judgments of interest, obviousness, and familiarity more than those of complexity or abstractness. The judges, however, were acquainted with the purpose of the study and with the fact that the difficulty of the tasks was known for children as well as for adults, and that the corresponding analysis was to be made. It would be a very difficult, if possible, task to obtain children's judgments of the several characteristics. There is no way of measuring the influence on the correlations computed of the effect of the adult viewpoint on the analysis made. It will be

TABLE IX

NUMBER AND TYPES OF ELEMENTS ON WHICH THE CORRELATIONS OF TABLE VII ARE BASED FOR EACH AGE AND GRADE GROUP

Group	P.C.*	P.R.*	G.C.*	G.A.*	G.S.*	Ar.*	V.O.*	S.C.*	Total
Mental Age Groups, 7–14 ...	10		20	10	24	20			84
Adult			20		24		140	22	206
Grades									
4							140	57	197
5							140	57	197
6	30		50	30	60	60	140	57	427
7			20		24		140	58	242
8	10		40	10	48	20	140	58	326
9							140	49	189
10							140	57	197
12	20		20	20	24	40		57	181
Chronological Age Groups, 7–14	10	6	10						26

* P.C. = Picture Completion
P.R. = Picture Reorganization
G.C. = Geometrical Completion
G.A. = Geometrical Analogy
G.S. = Geometrical Series
Ar. = Arithmetic Computation
V.O. = Verbal Opposites
S.C. = Sentence Completion

well to keep in mind the qualifying phrase "from an adult viewpoint" in interpreting the correlations presented here, as well as other considerations involving the judgments presented in Table IV.

In spite of the indicated differences in the relations between the various traits and difficulty, certain trends are apparent. These have already been mentioned above for all groups taken together. In Table X is given a further analysis for the three types of groups, disregarding the adult group.

The M.A. groups are definitely a subnormal group. That is, on the average they are appreciably below the C.A. and grade groups. The latter are unselected, the former have an average I.Q. of about .80.* It is interesting to note that with this subnormal group, complexity is the outstanding factor in causing difficulty, familiarity and obviousness being minor though appreciable factors. It is surprising that elaborateness is less conspicuous. The situations on which the correlations for the mental age groups

* Slawson, J., "Psychology of the Juvenile Delinquent" (unpublished thesis).

are based are all non-verbal in character. This restricts the possible variability in the traits. The variability in abstractness, symbolicness, and interest is much less than is true of the other traits. This is probably the reason for the low correlations of difficulty with the former three traits.

The test elements given to the C.A. groups were also all of the non-verbal type, while those given to the grade groups were a mixture of verbal and non-verbal. This means that the possibility of variation is greater in the latter case for all traits and particularly for abstractness, symbolicness and intrinsic interest. With this in mind, it may be concluded that there is not much difference between the nature of difficulty for the C.A. and grade groups. Disregarding the correlations for the M.A. groups, it is seen that obviousness and familiarity are, as with the superior adult group, the most potent factors in determining difficulty, complexity being not far behind. The differences are much less marked here than with the adult group, indicating a difference in degree rather than in kind. Interest is less potent with the younger groups, and symbolicness and elaborateness though less important are still not entirely negligible. If a greater diversity of material were used with the C.A. groups, it might well be that abstractness would also appear as a factor of importance.

The data need further analysis. Such situations should be used as will allow for a fairer comparison between traits. That is, the situations should be such that the variation in all traits is the same for each group as well as the variation in difficulty of the tasks. The types of elements should be uniform for all groups; there should be verbal as well as non-verbal material in every case.

A much more detailed analysis would undoubtedly be more enlightening and more practical. For example, what would be the correlations between difficulty and the various traits for any grade for each type of element? The amount of work involved makes such an analysis prohibitive for this study, but it would be worthwhile from the standpoint of practical application.

In addition to the above analysis, judgments on six traits, complexity, elaborateness, abstractness, familiarity, obviousness, and interest, were obtained for 95 arithmetic reasoning tasks and 95 completion tasks.[1] The actual difficulty of the elements was not

[1] This material was obtained from Professor Thorndike and was in the process of being standardized.

obtained, but they were arranged approximately in the order of their difficulty. The judges in this case were twelve graduate students of education in the summer session at the University of Kansas. Six judges worked with the arithmetic tasks and six with the completion tasks. The correlations between their judgments and the difficulty of the elements is given in Table XI. In this case the order of difficulty is not for any particular age or grade group. Although these correlations are for one type of material alone, the variation between the least and most in most of the traits was considerable. This was probably not true in abstractness, interest, and symbolicness. The judges here were not restricted in their judgments by having to deal with several types of elements, nor did they give ratings to the elements. They merely ranked them in order from least to most for each trait.

Table XI may be interpreted as meaning that in general all the traits considered are highly correlated with difficulty in the case of verbal completions; whereas in the case of arithmetic, abstractness and interest are much less potent factors than are complexity, elaborateness, familiarity, and obviousness.

TABLE X

MEDIAN CORRELATION BETWEEN DIFFICULTY AND EACH OF THE SEVEN CHARACTERISTICS, FOR AGE AND GRADE GROUPS

Groups	Characteristics						
	C*	E*	A*	S*	F*	O*	I*
Grades	+.625	+.213	+.524	−.308	−.706	−.698	+.579
Adults	+.487	+.042	+.529	−.057	−.626	−.800	+.636
Chronological Ages	+.558	+.466	−.082	−.376	−.645	−.538	−.390
Mental Ages	+.690	+.369	−.150	−.239	−.446	−.459	+.093

C = Complexity
E = Elaborateness
A = Abstractness
S = Symbolicness
F = Familiarity
O = Obviousness
I = Interest

TABLE XI

CORRELATIONS BETWEEN DIFFICULTY AND CERTAIN CHARACTERISTICS FOR ARITHMETIC PROBLEMS AND SENTENCE COMPLETIONS

Characteristics	Completions	Arithmetic
Complexity	+ .903	+ .860
Elaborateness	+ .781	+ .875
Abstractness	+ .720	+ .335
Familiarity	− .874	− .608
Obviousness	− .916	− .809
Interest	+ .629	− .358

CHAPTER V

DIFFICULTY AND INTELLIGENCE

Is difficulty the same for several groups of the same mental level but differing in chronological age, i.e., in intelligence quotients? The available data restrict the question to a comparison of subnormal with normal, or rather with unselected groups. The question has been partially answered for the nature of difficulty in the preceding chapter.

Table III allows for a different sort of answer. How will two groups of the same mental age—one group subnormal, the other normal—compare in their responses on the same tasks? Since the C.A. groups and the M.A. groups did not attempt the same tasks, the comparison must be between the M.A. groups and grade groups of about the same average age. Roughly, a mental age of 12 for this group of subnormals may be compared with a sixth grade group, since the average age of sixth-grade children is about 12 years.

Table XII is a repetition of such of the data of Table III as allow for this comparison. This comparison is available only for mental age 12 which corresponds to grade 6. To emphasize the comparison the difficulty values for mental age 11 are included.

Averages alone in this case are meaningless. A comparison of the differences between difficulty values would be more meaningful. A simpler comparison is to list the number of cases in which the mental age groups find tasks easier and the number in which they find them harder than the corresponding grade groups. In very few cases are there great differences in difficulty.

Table XIII, which is shown on page 30, gives such a summary for Table XII.

This approximate comparison indicates that a subnormal group can do tasks at least as well as a normal group of the same mental age. Probably they can do somewhat better.

Of the individuals composing the M.A. groups referred to in the previous chapter, those were chosen whose records on the Stanford Revision of the Binet Simon Intelligence Test were available, and

TABLE XII

DIFFICULTY VALUES OF CERTAIN ELEMENTS FOR GRADE 6, AND MENTAL-AGE GROUPS 11 AND 12

Test Elements		MA 11	MA 12	Gr 6
Thorndike Form IA				
Test 1	1	50	77	86
	2	34	54	71
	3	41	58	71
	4	59	79	81
	5	44	64	72
	6	56	76	71
	7	33	63	55
	8	39	60	53
	9	43	46	34
	10	33	52	35
	11	14	16	21
	12	23	35	28
Form IL	1	79	82	89
Test 1	2	73	88	89
	3	43	67	59
	4	44	63	55
	5	56	70	66
	6	58	67	65
	7	67	74	71
	8	63	79	80
	9	59	79	76
	10	42	61	60
	11	35	45	49
	12	29	38	45
Form IA	1	84	79	80
Test 2	2	68	61	55
	3	43	43	51
	4	56	59	57
	5	46	44	53
	6	38	41	34
	7	29	33	30
	8	12	11	13
	9	28	38	29
	10	12	18	8
Form IL	1	48	68	70
Test 2	2	63	71	61
	3	60	70	66
	4	58	63	59
	5	53	58	66
	6	45	57	43
	7	22	31	29
	8	15	25	27
	9	39	31	39
	10	11	17	17

Test Elements		MA 11	MA 12	Gr 6
Thorndike Form IIB				
Test 1	1	84	90	90
	2	32	36	53
	3	86	89	81
	4	16	17	11
	5	44	48	37
	6	69	72	50
	7	10	11	2
	8	26	29	8
	9	33	32	41
	10	6	7	1
Test 2	1	10	13	4
	2	19	22	14
	3	20	29	22
	4	11	15	14
	5	5	12	3
	6	1	3	2
	7	2	4	3
	8	10	22	8
	9	5	4	7
	10	24	31	15
Test 4	1	82	90	96
	2	81	89	98
	3	61	81	74
	4	45	63	29
	5	79	96	91
	6	78	92	89
	7	67	82	82
	8	25	46	30
	9	74	86	89
	10	73	89	86
	11	63	78	76
	12	27	56	20
	13	34	50	45
	14	60	69	51
	15	50	70	43
	16	13	33	13
	17	16	23	11
	18	6	15	3
	19	13	20	11
	20	4	8	3

TABLE XIII

SUMMARY OF TABLE XII, SHOWING COMPARISON OF DIFFICULTY FOR MENTAL AGES 11 AND 12 WITH GRADE 6

Of the 84 elements taken by	The number easier for the grade group is	The number about equally difficult for both groups is	The number more difficult for the grade group is
Grade 6 and Mental Age 11	44	24	16
Grade 6 and Mental Age 12	23	10	51

The average difficulty for the 84 elements for grade 6 is
6.25% higher than for mental age 11
2.5 % lower " " " " 12

whose I.Q.'s did not exceed 80.[1] This upper limit was chosen merely to obtain as large populations as possible without reaching too near the normal group.

From the individual records it was possible to determine the percentage of each mental age group that responded to each test correctly. Terman (1917) gives corresponding percents of correct response for the unselected groups on which the test was standardized in this country. A comparison of these percents on each test for corresponding M.A. and unselected C.A. groups is shown in Table XIV.[2] This table corroborates the findings given above. The subnormal groups average at least as well as the normal groups of the same age. The interpretation of this condition that is usually given is probably the right one: namely, that the longer number of years the subnormals have lived than the normals of the same mental age enables them to develop more, "horizontally," than those who are younger.

Hollingworth (1920) reports two experiments in learning that compare a subnormal group with a matched normal group, the two groups being of the same mental age. In both cases the subnormal group did as well as the normal group.

[1] The individual test records were made available through the courtesy of Dr. C. P. McCord, health director of the Albany, New York, Public Schools and psychiatrist for the Berkshire Industrial Farm at Canaan, New York; and of Colonel E. C. Barber, director of the House of Refuge, New York City.

[2] See footnote to page 9. Only a summary table for Table XIV is given here.

TABLE XIV

SUMMARY

Age	Difference between % correct for subnormal and normal groups	Number of tests of Stanford Binet in which		
		Subnormal group exceeds normal group	Normal group exceeds subnormal group	Normal and subnormal groups are equal
7	7	32	15	3
8	11	29	8	1
9	5	26	19	2
10	7	42	16	6
11	5	43	21	6
12	2	28	22	7

This table reads: Of those elements of the Stanford Binet test which were attempted by both an unselected group of chronological age 7 and a subnormal group of mental age 7, the average percent correct for the latter group was 7% greater than for the former group. In the case of 32 tests (of different age levels) the subnormal group had a higher percent of correct response; in the case of 15 tests the unselected group had a higher percent. and in the case of 3 tests the percent of correct response was the same for both; etc.

CHAPTER VI

LEVELS OF DIFFICULTY

The data used in the preceding chapters are of value in understanding the significance of age and grade status. From the data one can tell how well each age or grade group can do certain tasks. Or one can tell which situations are easy and which difficult for the different groups.

Table XV* shows, for each group, those tasks that are correctly performed by at least 75 percent of the group and those correctly performed by not more than 25 percent. Table XVI* gives similar data with 90 percent and 10 percent as the limiting values. Reference to the actual test elements will make clearer the meaning of the level of difficulty of, for example, nine years, or the sixth grade.

In so far as the difficulty values are inaccurate, are unreliable, differentiation within a small range is made meaningless. This is particularly true for percents close to zero, since the probable errors are comparatively large. In some instances, inaccuracies are apparent; for example, where the difficulty value of a test element for an age group is lower than the difficulty value for the same test element for a younger group. In such instances in the tables, the value for the younger group is taken as the correct one and the value for the older group is taken as of the same order, unless, for some reason, the former is obviously inaccurate.

No further analysis of the difficulty levels given in Tables XV and XVI is attempted. Such analysis here could be no more than an enumeration of the names of the tasks or a description of them. It is more satisfactory to study the elements themselves. It would be interesting to examine in detail the discrepancies between mental age, chronological age, and grade groups. This, however, would form a study in itself, and is beyond the scope of the present one.

In making the comparison of difficulty levels from Tables XV and XVI, it should be kept in mind that all groups did not attempt the same elements. The mental age groups can be compared,

* See footnote to page 9. Only part of each of these tables is shown here.

since they all attempted the same 84 elements. Similarly the chronological age groups all attempted the same 26 elements, and so may be compared with each other. But the mental age and chronological age groups cannot be compared with each other in this respect, since there is not a single element that they both attempted. The grade groups 4 to 10 may be compared only for the Wylie Opposites Tests and Trabue Completion-Test Language Scales. The sixth, seventh, and eighth grade groups may be compared with the mental age groups for such of the elements as are common to all. Within groups it may be interesting to compare the elements done correctly by more than 75 percent with those done by not more than 25 percent, and to examine those elements done by not more than 25 percent of a group that are done by 75 percent of an older group.

TABLE XV

(Only part of this table is shown here)

ELEMENTS CORRECTLY DONE BY (1) AT LEAST 75%, (2) NOT MORE THAN 25% OF THE GROUP, FOR EACH AGE AND GRADE GROUP

Group	Over 75%	Under 25%	Group	Over 75%	Under 25%	Group	Over 75%	Under 25%
Chron. Age			Mental Age			Grade		
7	P4–1	P4–6, 8–10 P5–3–10 P6–2–6	7		1A1–2, 3, 5, 7–10 1L1–3–12 1A2–3–10	4	Wb–1, 2 Wc–1, 7 Wd–1–6 B–1–3 C–1, 2	Wa–7–35 Wb–8–35 Wc–3–6, 9–35 Wd–8–35 B–7–10
8	P4–1	P4–6, 8–10 P5–5–10 P6–3–6			1L2–1, 4–10 2B1–2, 4, 5, 7–10		D–1, 3 E–3 F–2	C–6–10 D–7–9 E–7–9 F–7, 8
9	P4–1, 3, 7	P4–8, 9 P5–5–10 P6–3–6			2B2–1–10 2B4–4, 6–8, 11–20			J–4 K–3, 4 L–2–5 M–1, 2, 5, 7

P4 = Pintner Test 4 — 1A1 = Thorndike Form 1A Test 1 — Wb = Wylie Form b
P5 = " " 5 — 1L1 = " " 1L " 1 — Wc = " " c
P6 = " " 6 — 1A2 = " " 1A " 2 — Wd = " " d
etc. — etc.
B, C, D, etc. = Trabue Form B, C, D, etc.

The notation in the table has the following meaning: The numbers and letters before the first dash in each case represent the name of the test and the form. The numbers after the first dash represent the items of the test. Thus P4–6, 8–10 means Pintner Test 4, items 6 and 8 to 10.

TABLE XVI

(Only part of this table is shown here)

TEST ELEMENTS CORRECTLY DONE BY (1) AT LEAST 90%, (2) NOT MORE THAN 10% OF THE GROUP, FOR EACH AGE AND GRADE GROUP

Group	Over 90%	Under 10%	Group	Over 90%	Under 10%	Group	Over 90%	Under 10%
Chron. Age 7		P—6, 8, 9 P5—5–10 P6—4–6	Grade 5	Wb—1 Wc—1 Wd—3, 4 B—1–3 C—1, 2 D—1 F—2	Wa—7–19, 21–35 Wb—9,12–35 Wc—3,5,6, 9–35 Wd—9–19, 21–35 B—8, 10 C—8–11, 6 D—8, 9 E—8, 9 F—8 J—4 K—4 L—5, 4 M—5, 7	Grade 8	1G2—1 2B1—1 2B4—1, 2, 5–7,9–11, 14, 15 Wa—3 Wb—1–3, 5, 6 Wc—1, 7 Wd—1–6 B—1–4 C—1–4 D—1, 3 E—3, 4 F2, 3 J—1 K—1	LG2—10 2B1—7, 10 2B2—1,6,7 Wa—9, 15, 17,18,21–24, 26–35 Wb — 12, 13, 18–25 Wc—5,13–16,18–27, 29–35 Wd — 13, 16,18,19, 23–35 B—10 C—10 M—7
8	P4—1	P4—9 P5—5–10 P6—5, 6						
9	P4—1	P5—7–10 P6—5–6						
10	P4—1, 3	P5—7–10 P6—5, 6						
11	P4—1, 3	P5—7,9,10						
12	P4—1, 3, 7	P5—9, 10						
13	P4—1, 3, 7	P5—10						
14	P4—1, 3, 7	P5—10						

CHAPTER VII

SUMMARY AND CONCLUSIONS

The limitations of the study are apparent. The lack of control of certain factors makes the interpretation of results difficult, and lessens their value to some extent. Here may be included such factors as type of tasks—test elements as against more real situations; distribution of types for the several groups of subjects; number of tasks of each type; number of subjects; the time element in performing the tasks; the number of judgments for each trait; the viewpoint of the judges. These factors have been cared for in part, but the nature of the problem, as explained in the earlier chapters, makes it difficult to control them all adequately. The value of the study is dependent, therefore, on the extent to which these factors are influential. This has been discussed previously in some detail.

The data and discussions of the foregoing chapters indicate the following for such tasks as this study makes use of, under timed conditions:

1. Complexity of the process of getting the solution, obviousness of the solution, and familiarity of the subject matter are, in general, the most highly correlated with difficulty, the correlations being about .60.
2. Abstractness of the process of getting the solution, and the intrinsic interest of the task are, in general, uncorrelated with difficulty.
3. Elaborateness of the form of the task, and the symboliceness of the form of the task are correlated to a slight, though appreciable, degree with difficulty, the correlations being about .35.
4. For adults, obviousness and familiarity are the most important factors in determining difficulty. Complexity and abstractness are also fairly important factors. Elaborateness and symboliceness are negligible. The lack of validity of the judgments for interest makes the interpretations based on them of little value.

5. For subnormal children, complexity is the outstanding factor in causing difficulty.
6. For normal children, the relations of the several characteristics to difficulty are much the same as with adults, complexity being more important with them than with adults.
7. The validity of the judgments made is sufficiently high, with the exception of those for interest, to make conclusions based on them meaningful.

These statements are true for the various groups for all types of tasks used in this study, treated as a group and not for any particular type of subject matter or task.

As for the method used in making an investigation of difficulty the following may be pointed out:

1. A more accurate and more meaningful analysis could be made if the same test elements were used for each group.
2. A more even distribution of types of test elements should be used for each group. That is, about the same number of verbal tasks as of non-verbal tasks, of number tasks as of word tasks, of analogies as of completions, etc., should be employed.
3. Such data would allow for an analysis of difficulty for different types of subject matter on a comparable basis, as well as of the meaning of difficulty for different groups of individuals.

The construction of intelligence tests can be made much finer and more valuable by an analysis such as the one made in this investigation. A careful study of test elements from the viewpoint of what makes them difficult would aid materially in producing more precise and discriminating instruments. The same is true of more real situations, such as those employed in the classroom. The application of knowledge of the meaning of difficulty would provide a material aid to the teaching and learning process.

BIBLIOGRAPHY

Buckingham, B. R. Deriving and Publishing Evaluated Material. (Editorial). *Jl. Ed. Res.* 3: 139, 1921.

Hollingworth, L. S. *The Psychology of Subnormal Children.* 1920.

Kelley, T. L. *Statistical Method.* 1923.

Pintner, R. A Non-Language Group Intelligence Test. *Jl. App. Psych.* 3: 199, 1917.

Slawson, J. The Psychology of the Juvenile Delinquent. (Unpublished Doctor's thesis, Columbia University.)

Terman, L. M. *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence.* 1917.

Thorndike, E. L. A Non-Verbal Group Intelligence Test. *Jl. App. Psych.*, 1919.

Trabue, M. R. *Completion-Test Language Scales.* 1916.

Wylie, A. T. *The Opposites Tests.* 1925.

List of Articles Giving Objective Data on Difficulty of Test Elements or Other Tasks

I. Intelligence Tests

Healy, Wm. Pictorial Completion Test. *Jl. App. Psych.* 5: 225.

Irwin, H. N. A Preliminary Attempt to Devise a Test of the Ability of High School Pupils with Mental Manipulation of Space Relations.

Lowell, F. A Group Intelligence Scale for Primary Grades. *Jl. App. Psych.* 3: 215.

Pintner, R. A Non-Language Group Intelligence Test. *Jl. App. Psych.* 3: 199.

Pintner, R. and Toops, H. A. A Drawing Completion Test. *Jl. App. Psych.* 2: 164.

Town, C. A Mass Verbal Test for Use with Kindergarten and First Grade. *Jl. App. Psych.* 6: 89.

II. Mathematics and Science

Chapman, J. C. The Measurement of Physics Information. *Sch. Rev.* 27: 748.

Harris, E. and Breed, F. S. Comparative Validity of Hotz and Rugg-Clark Tests. *Jl. Ed. Res.* 6: 393.

Jones, F. T. Practice Exercises in Physics. *Sch. Rev.* 26: 341.

Minnick, J. A. A Scale for Measuring Pupils' Ability in Geometry. *Ed. Rev.* 27: 103.

Powers, S. R. A Comparison of the Achievement of High School and University Students in Certain Tasks in Chemistry. *Jl. Ed. Res.* 6: 332.

Randall, D. P., Chapman, J. C., and Sutton, C. The Place of Numerical Problems in High School Physics. *Sch. Rev.* 26: 42.

Rugg, H. O. Experimental Determination of Standards in First Year Algebra. *Sch. Rev.* 24: 37.

———. Standardized Tests and the Improvement of Teaching in First Year Algebra. *Sch. Rev.* 25: 122.

III. History and Geography

Barthelmess, H. Geography Testing in Boston. *Jl. Ed. Res.* 2: 701.

Brooks, S. S. Getting Teachers to Feel the Need for Standardized Tests. *Jl. Ed. Res.* 2: 425.

Scott, F. and Myers, G. Children's Empty and Erroneous Concepts of the Commonplace. *Jl. Ed. Res.* 8: 327.

Stevenson, P. R. A Problem Test in Geography. *Jl. Ed. Res.* 5: 350.

IV. Language

Ashbaugh, E. K. The Measurement of Language. *Jl. Ed. Res.* 4: 32.

Charters, W. W. Constructing a Language and Grammar Scale. *Jl. Ed. Res.* 1: 249.

Henmon, V. A. C. Standardized Vocabulary and Sentence Tests in French. *Jl. Ed. Res.* 3: 81.

Wilson, G. Language Error Tests. *Jl. Ed. Psych.* 13: 341.

Gates, A. I. A Test of Ability in Pronunciation of Words. *Teachers College Record.* 26: 205.

V. Miscellaneous

Chapman, J. C. and Toops, H. A. A Written Trade Test. *Jl. App. Psych.* 3: 358.

VITA

Jacob Samuel Orleans was born in London, England, on December 22, 1899. He migrated to this country in April, 1905. He attended the public elementary and high schools of New York City, graduating from the De Witt Clinton High School of that city in June 1916. He graduated from the College of the City of New York in June 1919 with the B.S. degree. He held the position of statistician for the Times Square Auto Supply Co. of New York City from January to August 1920. September 1920 he registered for graduate study in Educational Psychology at Teachers College, Columbia University, and pursued his studies until June 1924. While doing his graduate work at Teachers College, he obtained the degree of M.A., October 1921; held the position of Research Assistant in the Institute of Educational Research of Teachers College during the years 1921–1922 and 1923–1924; taught mathematics in the New York City High Schools, 1921, and 1922–1923; instructed in educational psychology and measurements, and in high school teaching during the summer sessions 1922 and 1923 at Cornell University, and 1924 at the University of Kansas; offered an extension course at the College of the City of New York, in the Teaching of Junior High School Mathematics. At present, he is Research Associate in the Bureau of Educational Measurements of the New York State Department of Education.